ASTROINFORMATICS
Introduction to AstroStatistics and Big Data in Astronomy

Darko Jevremović

Astronomical Observatory Belgrade, Fall 2018

Location and Time: Tuesdays 17:00-19:00, **AOB Library**

Office Hours: Any time when my office door is open;
After class is the best - ping me via e-mail.

Grading: 8-10 homeworks, 80% ; final exam: 20%
key: >85%=10, >75%=9, >65%=8, >55%=7, >45%=6 .

Class web site: https://github.com/astromundus/astroinfo2018

Reference textbook:
Ivezić, Connolly, VanderPlas & Gray: *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data* (Princeton University Press, 2014)
See http://press.princeton.edu/titles/10159.html

Learning Goals:

This course will introduce students to most common statistical and computer science methods used in astronomy and other physical sciences. It will combine theoretical background with examples of data analysis based on modern astronomical datasets. Practical data analysis will be done using python tools, with emphasis on astroML module (see www.astroML.org). While focused on astronomy, this course should be useful to all students interested in data analysis in physical sciences and engineering. The lectures will be aimed at undergraduate students and the main discussion topics will be based on Chapters 4 and 5, and selected topics from Chapters 6-10, from the reference textbook.

By taking this course, students will develop basic understanding of topics such as robust statistics, hypothesis testing, maximum likelihood analysis, Bayesian statistics, model parameter estimation, the goodness of fit and model selection, density estimation and clustering, unsupervised and supervised classification, dimensionality reduction, regression and time series analysis. Most of these topics will be applied in class homeworks to analysis of astronomical data.

Prerequisites: The students taking this class are required to have basic calculus and basic python skills, as well as basic scientific measurements and statistics skills at the level of a freshman lab.

Lecture format: New material will typically be covered during the first part of the class, while the second part will be more focused on practical data analysis work.

Class Schedule:

- WEEK 1 : Introduction to class: syllabus, literature, astroML, python, matplotlib, GitHub.
- WEEK 2 : Introduction to statistics (probability, distributions, robust statistics, Central Limit Theorem, hypothesis testing).
- WEEK 3 : Maximum likelihood and applications in astronomy.
- WEEK 4 : Bayesian statistics and introduction to MCMC.
- WEEK 5 : Model parameter estimation and model selection.
- WEEK 6 : Time series analysis.
- WEEK 7 : Dimensionality reduction and regression.
- WEEK 8 : Density estimation and clustering.
- WEEK 9 : Unsupervised and supervised classification.
- WEEK 10 : Review
- FINAL EXAM

Homework

There will be eight to ten homeworks, assigned at the end of session on Tuesday, and due next Tuesday. They will be centered on practical work using python and designed to test the weekly progress.

We will use modern software engineering tools, GitHub and Jupyter notebooks for HW submission.